

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 05-151253

(43)Date of publication of application : 18.06.1993

(51)Int.Cl. G06F 15/38
G06F 15/40
G06F 15/40

(21)Application number : 03-315786

(71)Applicant : CANON INC

(22)Date of filing : 29.11.1991

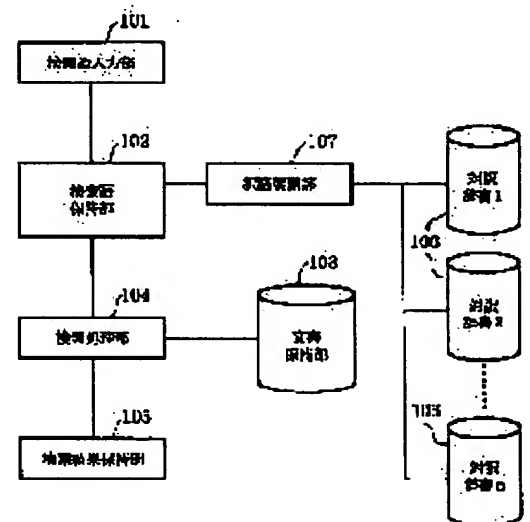
(72)Inventor : UEDA TAKANARI
ITO SHIRO
FUJITA MINORU

(54) DOCUMENT RETRIEVING DEVICE

(57)Abstract:

PURPOSE: To provide a document retrieving device capable of retrieving a document by a language different from the language of an inputted retrieving word.

CONSTITUTION: A retrieving word inputted from a retrieving word input part 101 is stored in a retrieving word storing part 102. A translated word developing part 107 retrieves corresponding translation dictionaries 106 by means of the inputted retrieving word and adds translated words corresponding to the inputted retrieving word based upon various languages to the storing part 102 as retrieving words. A retrieving processing part 104 retrieves a document including the retrieving words stored in the storing part 102 from documents stored in a document storing part 103 and stores the document extracted as the retrieved result in a retrieved result storing part 105.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁(JP)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開平5-151253

(43)公開日 平成5年(1993)6月18日

(51)Int.Cl. ⁵	職別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 15/38		C 9194-5L		
15/40	5 0 0 Q	7060-5L		
	5 1 0 M	7060-5L		

審査請求 未請求 請求項の数1(全 5 頁)

(21)出願番号 特願平3-315786

(22)出願日 平成3年(1991)11月29日

(71)出願人 000001007

キャノン株式会社

東京都大田区下丸子3丁目30番2号

(72)発明者 上田 隆也

東京都大田区下丸子3丁目30番2号 キャ
ノン株式会社内

(72)発明者 伊藤 史朗

東京都大田区下丸子3丁目30番2号 キャ
ノン株式会社内

(72)発明者 藤田 稔

東京都大田区下丸子3丁目30番2号 キャ
ノン株式会社内

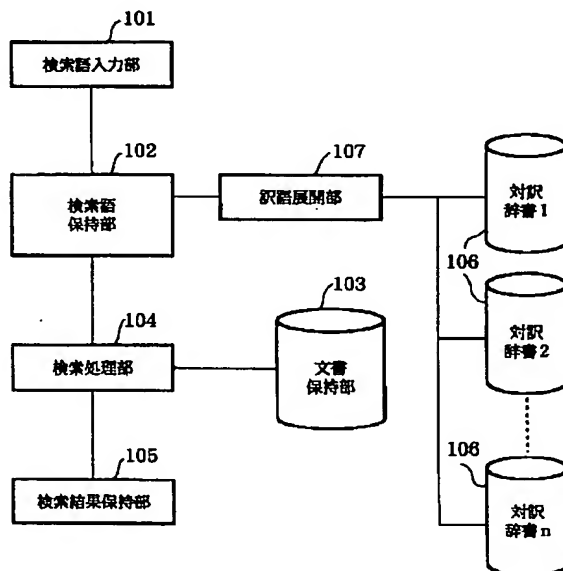
(74)代理人 弁理士 大塚 康德 (外1名)

(54)【発明の名称】 文書検索装置

(57)【要約】

【目的】 入力された検索語とは異なる種類の言語による文書の検索をも可能とし、使い勝手の良い文書検索装置を提供する。

【構成】 検索語入力部101より入力された検索語を検索語保持部102にて保持する。訳語展開部107では、この入力された検索語により、対訳辞書106を検索して、入力された検索語に対応する各種言語による訳語を検索語として検索語保持部102に追加する。検索処理部104は検索語保持部102に保持された検索語を含む文書を文書保持部103に保持されている文書の中から検索し、この検索の結果抽出された文書を検索結果保持部105に保持する。



【特許請求の範囲】

【請求項1】 検索語を入力して、前記検索語に基づいて文書保持部に保持されている文書を検索する文書検索装置であって、

各種言語間の対訳を登録してある対訳辞書と、
入力された検索語の各種言語における訳語を前記対訳辞書から検索する訳語検索手段と、
前記入力された検索語と前記訳語検索手段により検索された訳語とを検索語として文書を検索する文書検索手段とを備えることを特徴とする文書検索装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、計算機によって文書の検索を行う文書検索装置に関する。

【0002】

【従来の技術】文書データベースの普及と、計算機処理能力の向上により、大量の文書データベースから文書を検索する文書検索装置が広く用いられるようになってきた。

【0003】特に最近では、検索の際の柔軟性を向上させるために、予め文書にキーワードを与えず、ユーザが自由にキーワード（以下、検索語と称する）を指定できるようにした全文検索の方式が用いられるようになってきている。

【0004】

【発明が解決しようとする課題】ところで、文書データベースに保存される文書は種類の言語に限るものではない。例えば、日本語、英語、フランス語、ドイツ語等の言語によって記述されている文書が文書データベース内に混在している。このような場合、予め文書にキーワードを与えておき、このキーワードにて検索するキーワード方式のシステムでは、キーワードを一つの言語に統一すれば良い。しかしながら全文検索の方式による従来の文書検索装置では検索語として与えた言語以外の言語の文書は検索できない。例えば、検索語を日本語で与えれば日本語の文書しか検索できない。そして、他の言語による文書を検索したい場合は、それぞれの言語による検索語を別々に入力しなければならないので検索に手間がかかるという問題があった。

【0005】本発明は上記の問題点を鑑みてなされたものであり、全文検索方式においても、キーワード方式においても、入力された検索語とは異なる種類の言語による文書の検索をも可能とし、使い勝手の良い文書検索装置を提供することを目的とする。

【0006】

【課題を解決するための手段】上記の目的を達成するために本発明による文書検索装置は以下の構成を備える。即ち、検索語を入力して、前記検索語に基づいて文書保持部に保持されている文書を検索する文書検索装置であって、各種言語間の対訳を登録してある対訳辞書と、入

力された検索語の各種言語における訳語を前記対訳辞書から検索する訳語検索手段と、前記入力された検索語と前記訳語検索手段により検索された訳語とを検索語として文書を検索する文書検索手段とを備える。

【0007】尚、特許請求の範囲で言う「文書検索手段」による文書検索の方式は、全文検索方式、キーワード方式のいずれであっても良い。

【0008】

【作用】上記の構成により、使用者がある言語により検索語を入力し文書の検索を実行すると、検索語に対する各種言語への翻訳が実施され、入力された検索語とこれに対する各種言語に翻訳された検索語とを用いて各種言語の文書の検索を実行する。

【0009】

【実施例】以下に添付の図面を参照して本発明の好適な実施例について詳細に説明する。

【0010】<実施例1>実施例1では検索語が日本語で入力される場合を例にとって説明する。

【0011】図1は、実施例1による装置の構成を示すブロック図である。同図において、101は検索する語を入力する検索語入力部である。102は検索語保持部であり検索語入力部101から入力された検索語を保持する。103は文書保持部であり検索の対象となる複数の文書が保存されている。104は検索処理部であり検索語保持部102に保持された検索語を含む文書を文書保持部103に保存されている文書の中から検索する。105は検索結果保持部であり検索処理部104で検索した結果、抽出された文書を保持する。

【0012】106は対訳辞書であり、それぞれの単語に対する訳語を保持している辞書である。これは例えば対訳辞書1は日本語から英語への対訳辞書、対訳辞書2は日本語からドイツ語への対訳辞書というように、各言語ごとに用意されている。107は訳語展開部であり検索語保持部102に保持された検索語に対する訳語を求める。

【0013】図2は図1に示した装置における動作の処理手順を示すフローチャートである。本図を参照しながら実施例1の動作を説明する。

【0014】まず、ステップS1では検索語入力部101で検索語の入力を行う。入力された検索語は検索語保持部102に保持される。次にステップS2に移り、訳語展開部107において、検索語保持部102に保持された検索語の訳語を検索する。この検索は対訳辞書106中のすべての辞書（対訳辞書1から対訳辞書n）について実施する。

【0015】次にステップS3において、検索語の訳語があったかどうかを調べる。検索語の訳語がある場合は、ステップS4に移り、得られた全ての訳語を検索語保持部102に保持されている検索語に加える。このとき上述の訳語は「入力された検索語もしくは訳語（を含

む文書)」というOR条件の形で追加される。検索語の訳語が1つもない場合はステップS4をとばしステップS5に移り、検索語保持部102に保持される検索語は「入力された検索語のみ」ということになる。

【0016】次にステップS5に進み、検索処理部104において、検索保持部102に保持された検索語を含む文書を文書保持部103から検索する検索処理を行う。ここでの検索処理は、例えば複数の文字列を同時に検索するAC法を用いることにより実施される。検索処理の結果得られた文書は検索結果保持部105に保持する。

【0017】次に検索語入力部101において、検索語「コンピュータ」が入力された場合を例にとり、各部の説明を行う。なお、この例では文書保持部103には日本語文書と英語文書だけが保持されているものとする。

【0018】検索語「コンピュータ」を訳語展開部107で展開した結果、訳語として英語の"computer"がえられる。そこでこれを検索語保持部102にORの形で追加して、"コンピュータ"もしくは"computer"を含む文書を検索する。これによって日本語文書で"コンピュータ"を含む文書だけでなく、英語文書で"computer"を含む文書も検索できる。なお、フランス語やドイツ語等の文書が文書保持部103に保持されている場合もそれぞれの対訳辞書を備えることにより上記の例と同様にして検索すればよい。

【0019】＜実施例2＞上記実施例1では検索語として1つの単語を用いたが、検索語が複数の単語であっても、個々の単語に対して訳語展開の処理をすれば、実施例1と同様な方法で対応できる。この際、入力された検索語間の条件を、訳語展開後の検索語群の間でも保存するようにする。例えば、"ワークステーション" AND "パソコン" という形で検索後が入力された場合は、検索語間のAND関係を保存して、("ワークステーション" OR "work station") AND ("パソコン" OR "personal computer")とし、文書検索を実施すれば良い。

【0020】＜実施例3＞上記実施例では、訳語を全て検索語に加えて検索に用いたが、文書ごとにその言語によって検索語を変えてもよい。すなわち、常に"computer" OR "コンピュータ"を検索するのでなく、英語の文書の場合は英語の"computer"を検索語とし、日本語の文書の場合は日本語の"コンピュータ"を検索語とする。これは、各文書の言語を判定する手段と、検索語ごとにその言語の種類を保持する手段を設ければ実現できる。

【0021】各文書の言語を判定する手段の1例としては、文書から単語等を抽出し、その単語を見出し語として有する対訳辞書を検索することで言語の判定が可能となる。また、各文書毎にあらかじめ使用言語情報をもたせておいても良い。

【0022】入力された検索語の言語の種類についても上述の文書の言語の判定方法によって判定が可能である。さらに、対訳辞書により翻訳された検索語については、それぞれ使用した対訳辞書を検索語に対応させて記憶しておけばよい。

【0023】図3に実施例3による文書検索のフローチャートを示す。ステップS11にて検索語が検索語入力部101より入力される。次にステップS12において入力された検索語の種類を上述の方法により判定する。ステップS13にて、検索語に対する訳語を対訳辞書106より検索し、続くステップS14において、各検索語をそれぞれの言語の種類と対応させて検索語保持部102に保持する。

【0024】ステップS15では上述の方法により対象文書の言語を判定する。次にステップS16において、検索語保持部102よりその対象文書にて使用されている言語と同じ言語による検索語を抽出して、文書内にその検索語が存在するかどうかをAC法等により調べる。

【0025】そして、ステップS17にて対象文書に検索語が存在すればステップS18へ進み、検索結果として対象文書を検索結果保持部105に保持する。また、ステップS19において、未チェックの対象文書が存在するかどうかをチェックし、存在するのであればステップS15へ戻り、次の対象文書について上述の処理を繰り返す。また、未チェックの対象文書が存在しないのであれば、本処理を終了する。

【0026】尚、上記実施例1及び2では、検索語が日本語で入力される場合を例にしたが、他の言語であっても、それに対応する対訳辞書を用意することによって、同様に実施できることは言うまでもない。検索語がどの言語で入力されるかわからないような場合も、各種の言語間の対訳辞書を用意し、上述の実施例3で説明したような「入力された検索語の言語の種類を判定する手段」を追加すれば、同様に実施できる。つまり、入力された検索語の言語に対応する対訳辞書により入力された検索語に対する翻訳語を抽出し、検索語保持部102にOR条件で追加し、AC法等により文書検索を実行すればよい。

【0027】更に、上記の各実施例では、検索の際にAC法を用いたが、その他のパターンマッチング・アルゴリズムを用いてもかまわない。

【0028】更に、上述の各実施例では全文検索方式の場合について説明したが、文書にあらかじめキーワードが与えられているキーワード方式の場合についても上記各実施例と同様にして実施が可能である。

【0029】以上説明したように、上記の各実施例によれば、ユーザが指定した検索語の言語と異なる文書も検索できるので、利用しやすい文書検索装置がえられるという利点がある。

【0030】尚、本発明は、複数の機器から構成される

システムに適用しても1つの機器からなる装置に適用しても良い。また、本発明はシステム或いは装置に本発明により規定される処理を実行させるプログラムを供給することによって達成される場合にも適用できることはいうまでもない。

【0031】

【発明の効果】以上説明したように、本発明によれば、1種類の言語による検索語の入力から複数の種類の言語による文書の検索を可能とし、使い勝手の良い文書検索装置を提供することができる。

【0032】

【図面の簡単な説明】

【図1】本実施例による装置の構成を示すブロック図で*

*ある。

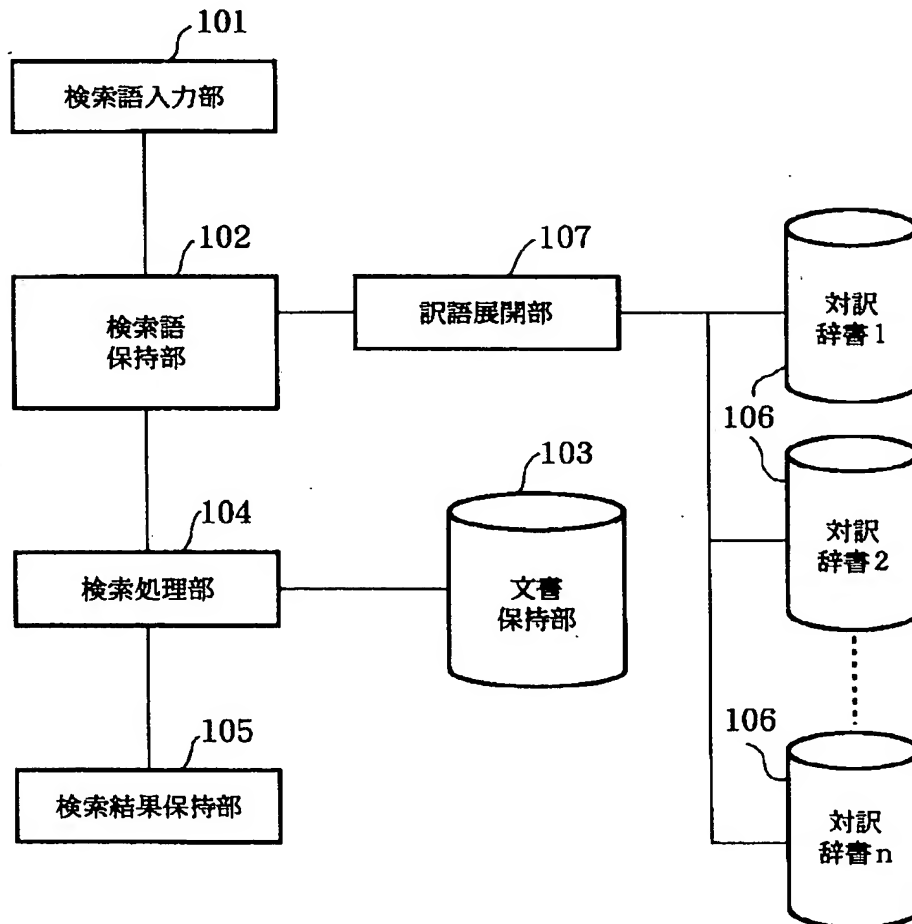
【図2】実施例1による文書検索処理手順を示すフローチャートである。

【図3】実施例3による文書検索処理手順を示すフローチャートである。

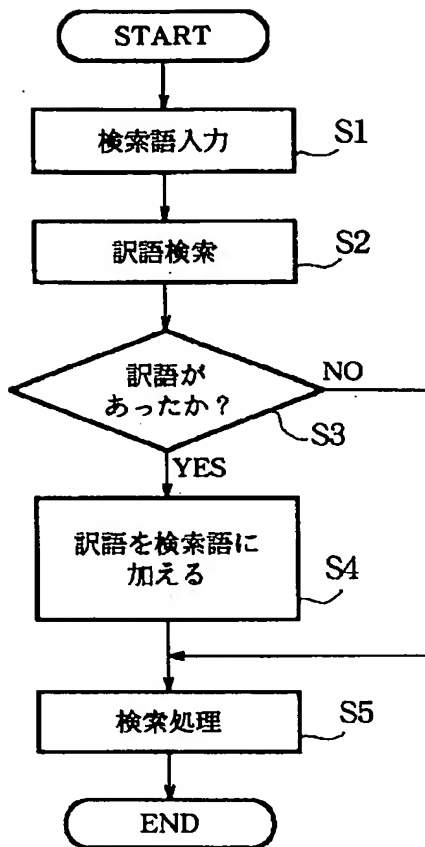
【符号の説明】

101 検索語入力部
102 検索語保持部
103 文書保持部
104 検索処理部
105 検索結果保持部
106 対訳辞書
107 訳語展開部

【図1】



【図2】



【図3】

